EVALUATION OF SINGLE-PLATFORM TECHNOLOGIES FOR ABSOLUTE CD4+ AND CD8+ CELLS

Maurice O'Gorman, Ph.D.

Northwestern University Children's Memorial Hospital Chicago, IL

INTRODUCTION

The CD4+ T cell count remains the most often ordered immunological procedure, and it is still our most valuable immunological marker in the evaluation of patients infected with HIV 1. A CD4+ count is an involved process, and there have been many milestones in the development of accurate and precise CD4+ counting since we began enumerating CD4+ T cells clinically in the late 70's to early 80's. These milestones have included direct conjugation of monoclonal antibodies, the ability to analyze three and four colors simultaneously, and the ability to measure CD4+ counts directly in whole blood. The flow cytometer allows for an accurate measurement of the percent lymphocytes expressing specific T cell antigens. Until recently, we were unable to obtain absolute T cell counts directly from the flow cytometer. We were dependent upon hematology in that we required a CBC and WBC differential to obtain absolute lymphocyte counts. Currently, we are entering a new era where we can finally get absolute counts off of a single platform. This presentation addresses three separate evaluations of single platform technologies for the measurement of absolute CD4+ counts. Throughout the manuscript I will allude to "the predicate method," which is actually three separate procedures. The first is flow cytometry which generates the percent T helper cells by measuring the percentage of lymphocytes expressing the CD3 and CD4 antigens using fluorochrome conjugated monoclonal antibodies. To get the absolute number of T cells per cubic milliliter of whole blood, a second procedure is required

which is a complete blood count, i.e. a CBC (for total white blood cells), and a third procedure which is the differential blood count (for percentage of total lymphocytes). The product of the WBC (white blood cells/mm3), the differential (% lymphocytes), and percentage of CD4+ T lymphocytes is the absolute CD4+ T cell count/mm3. These procedures in toto represent the "predicate method" for obtaining absolute CD4+ T cell counts, and is the most common process used in 1998. The requirement for hematology procedures represents both a practical problem as well as a source of bias and imprecision in the CD4+ count. In this manuscript we will discuss 3 technologies which generate an absolute CD4+ count from a single platform.

Each of these studies was designed by the New Technologies Evaluation Group (NTEG) which is a subcommittee of the Flow Advisory Committee (FAC), a committee under the umbrella of the NIAID. The FAC is charged with ensuring that flow cytometry procedures are done accurately and precisely, and the NTEG is specifically responsible for evaluating new flow cytometry based technologies which are market ready and of potential usefulness in a clinical trial setting.

The NTEG is chaired by myself and is composed of statisticians, clinical scientists, and technologists from across the US and Canada.* We have now completed or are in the process of completing the three different single platform CD4+ count technology evaluations, and herein summarize our results to date. All of the stud-

ies were performed in flow cytometry laboratories enrolled in the NIAID flow cytometry quality assessment program, who were all in good standing during the studies. The predicate method was subject to stringent quality control (QC) criteria, and any sample which failed the QC requirements was eliminated from the data base.

The first study was performed on a completely automated instrument, the **Imagn®** (Imagn) that is not a flow cytometer. There is no flow involved except the flow of the sample into chambers. The actual enumeration of individual cells is done on a static slide. The technology is volumetric capillary cytometry, and is discussed in more detail below. The second study was an evaluation of the Becton Dickinson True Count® system and was chaired by Dr. Bick. The third study was chaired by Dr. Keith Reiman and was an evaluation of both the Flow Count® and TetraOne® systems produced by Beckman Coulter. The latter two studies are both flow cytometry based and involve a ratiometric bead counting system for the absolute count measurement directly from the flow cytometer. For each study described, there were five test sites per study, and each individual site evaluated samples which were shipped from a central site as well as samples which had been procured locally.

SINGLE PLATFORM TECHNOLOGIES

Figure 1a is a representation of the Biometric Imagn® 2000 and the 4T8® cartridge. This is a fully automated device; the operator pipettes blood into cartridges, programs in the patient ID, and starts the instrument. The machine completes the rest of the test. The cartridges are designed to measure both CD4+ and CD8+ T cells, however only data on absolute CD4 counts is presented. The device goes through the following procedures to generate absolute counts. As seen in Figures 1a and 1b, samples are added into the loading chambers of individual, disposable, self-contained cartridges, (the machine can

hold 10 at a time). These loading chambers contain dried fluorochrome conjugated monoclonal CD3 and CD4 antibodies which are resuspended following the addition of the blood sample.

The cartridges sit on a rotating stage. Once all of the patient samples have been added, the door is closed and the stage shakes for about two minutes to get the antibody dissolved. The stage then does a fast spin, loading the samples into an incubation/mixing chamber for a twenty minute incubation where the labeling of the lymphocytes occurs. Each cartridge is then rotated to be positioned under a small piston which comes down on the cartridge and crushes a glass ampule diluting the stained blood sample. The chamber contains stainless steel mixing balls which are controlled with a magnet and used to mix the sample. After each cartridge has passed through the incubation/mixing stations the stage spins rapidly and forces the samples into finely calibrated capillary tubes. All this works by centripetal force, forcing the blood along at each successive stage. Within the capillary tubes, an exact volume of blood is scanned 10,000 times by a small red laser and the coincident fluorescence emission peaks are counted. Coincident emission peaks occur where CD4+ T cells are located due to their labeling with both CD3 and CD4 fluorochrome conjugated antibodies. The number of coincident peaks counted in a very precise volume is then multiplied by the dilution factor to generate the number of CD4+ T cells per &L of whole blood.

The other two technologies are flow cytometry-based technologies and utilize what is referred to as a ratio-metric method of absolute counting. Figure 2 illustrates the ratio-metric method of absolute counting with the Becton Dickinson True Count® system. In the first scattergram on the left, the X axis is CD45-PerCP (commonly referred to as a "lymphocyte gating reagent" as lymphocytes express characteristically high levels of CD45), and on the Y axis is right angle light scatter. This combination of parameters

allows for the identification of individual subpopulations or a three part differential of lymphocytes, monocytes and granulocytes. As flow cytometrists we are used to gating, so we draw an electronic gate around the lymphocyte subpopulation and then look at other parameters on those cells. In the second dot plot we are looking at CD3-FITC (fluorescence-1) versus CD4-PE (fluorescence-2) and you note the number of events which are both CD3+ and CD4+ (T helper cells). In a third dot plot which is ungated and displaying fluorescence-1 versus fluorescence-2, you gate on the bead population and note the number of beads counted in the sample. There are a known number of beads in the tube (provided by the manufacturer), and you have added a known volume of blood. In this example, there are 51,700 beads and you have added 50 &L of blood, therefore you know the bead concentration per &L of whole blood. After running the sample you know exactly how many beads were run and how many CD4+ T cells were run. The absolute CD4+ T cell count is therefore the ratio of the number of CD4+ T cells counted to the number of beads counted, times the number of beads per &L. This technology is practical, as you don't have to buy any new equipment. You're adding known numbers of beads to a tube to allow you to get absolute counting on a single platform instrument, i.e. This eliminates the your flow cytometer. requirement for hematological analysis.

The Beckman/Coulter single platform absolute count systems work essentially the same way with minor deviations. First, the beads in a known concentration are added to the tubes. For the second step, Beckman Coulter offers two different systems that have been FDA approved for absolute CD4+ counting. We have evaluated both of them. Figure 3 represents a summary of the Beckman Coulter Flow count® system as we evaluated it. This version is based on lymphocyte gating using only light scatter parameters. Forward angle light scatter is a relative correlate of the size of white blood cells and side scatter,

or right angle light scatter is a relative correlate of the granularity of the cells, and when you look at these parameters simultaneously the lymphocytes form a discrete cell cluster. An electronic analysis gate is drawn around the lymphocytes (first dot plot) and, you then measure the number of lymphocytes which are expressing both the CD3 and CD4 antigens within this gate in the second dot plot. In an ungated histogram (third dot plot) of this sample you obtain the total number of fluorospheres® counted. Since the concentration of the fluorospheres is known, the absolute count in the sample is equal to the ratio of CD4+ cells counted to the number of fluorospheres counted, multiplied by the known concentration of fluorospheres. The TetraOne® system is very similar to the BD method described above for lymphocyte gating except that the sample contains 4 monoclonial antibodies each conjugated with a different fluorochrome allowing you to gate lymphocytes with CD45, and measure absolute CD4+ and CD8+ T cells in the same tube.

PROTOCOL

Basically, we wanted to know how the results of the single platform technologies compared to our predicate method. What can you assess? You can measure accuracy, precision and "temporal fortitude," which is a measure of how well these assays stand up with time.

"Accuracy" as defined in the Webster's New Collegiate Dictionary is, "the closeness of the expected value to the true value of the measured quantity. It can be a measure of bias and it is sometimes called validity." What we have done is in fact to validate these methods. There is a problem with this definition of accuracy however because I do not believe (and most people would agree) that the predicate method is giving us the "true value" for CD4+ counts. So, within the NTEG we have had heated debates on whether or not we should call our assessments of these technologies "accuracy" because we are compar-

ing the new methods to the predicate method and the latter is not necessarily giving the true value. So, what do we do? Our solution is to measure how close we are with our new method to the predicate method and use the term "agreement" instead of accuracy when challenged.

Very simply, precision is the extent to which a measurement gives you the same results when repeated under identical conditions, and this is often referred to as reliability or variability. You can measure precision within a laboratory, in other words, how often can you get the same result on the same sample with the same method in your own laboratory? You can also measure between laboratory variability. If you are involved with clinical trials, you may be interested in variability between laboratories. Between laboratory variability includes additional variables not present in your own lab including different flow cytometers, different reagents, different operators and different hematology analyzers. As we know from previous experience, the hematology instruments are based on different technologies and they can give different answers, not so much with the white blood cell count, but definitely with the respect to the WBC differential.

In each of the studies, we measured accuracy or agreement as the difference in the CD4+ counts obtained on the same sample by the two different methods. For each study this was done on shipped samples as well as samples which had been procured locally. Each specimen had the CD4+ count enumerated by the predicate method and by the new technology. To measure accuracy, we subtracted the CD4+ count obtained by one technology from the CD4+ count obtained by the other technology. If both technologies generated the same results, the null hypothesis would state that the median difference between the 2 results is not going to be significantly different than zero.

What were our endpoints? We measured accuracy in CD4+ counting on samples procured locally and on shipped samples. Shipping adds several variables including temperature changes, physical abuse to the specimens and time. In our experience it takes 24-36 hours to get a shipped specimen into the laboratory. We also measured CD4+ and CD8+ counts by both methods on local samples when they were fresh, and again after the samples had been held for 24 hours. Comparison of the results between fresh vs. held specimens allows for an assessment of the effect of aging on CD4+ enumeration within each of the technologies, i.e. "temporal fortitude".

RESULTS

Although several endpoints were assessed, I will present only a summary of the most significant observations in our evaluations. The Biometric Imagn® evaluation is completed and has been published (O'Gorman et. al. CDLI, 4:173-179). The Becton Dickinson study is completed and the manuscript is in preparation. The Beckman/Coulter evaluation is not yet complete; the data is being tabulated and only preliminary observations are presented here.

Accuracy assessment is presented as the median, the tenth, and the ninetieth percentiles of the differences between the new technology and the predicate method. Table 1 is a summary of the differences between the Imagn® results and the predicate method (PM) results (CD4+ by PM minus CD4 by Imagn®) for all samples (n=570) procured locally in each of the laboratories. Overall, the PM generated higher CD4+ values than the Imagn®, with the median of the differences being 14 more CD4+ cells per &L, and ranging from 13 CD4+ cells (10th percentile) less than the Imagn® to 88 CD4+ cells (80th percentile) greater than the Imagn®. increase over the Imagn® was significant for the CD4+ counts in each of the three CD4+ groups, i.e., <200, 200-500 and >500. In addition, we observed that as the CD4+ count values

increased, the difference between the predicate method and the new technology also increased. This is referred to as a systematic bias. Table 2 is a summary of accuracy in each of the 5 laboratories, and similar to what was observed overall, each individual laboratory's predicate method generated higher CD4+ results than the results generated by the Imagn®. There were differences in each of the laboratories in the magnitude of the differences between results. This was determined predominately by the type of hematology instrument that was used to generate the absolute total lymphocyte count. This conclusion was reached because when the CD4+ percentages alone (without the absolute counts) were compared between the laboratories there were no significant differences.

Table 3 is a summary of the accuracy of the Becton-Dickinson True Count® (BD-TC) system. As alluded to above, this is a flow cytometry based ratio-metric method. Overall, the median of the differences between the two methods on 411 shipped samples (BD-TC CD4 minus PM CD4) was only seven CD4+ cells per &L with a range of 67 cells less (10th percentile) than the PM to 79 cells (80th percentile) greater than the predicate method. The difference is statistically significant, but it is fairly well balanced. In 2 groups, < 200 CD4+ cells and >500 CD4+ cells, there was no significant difference between these two methods. In the 200-500 CD4+ group, the difference is significant but again, it is balanced. However, when examining the differences between the methods in each individual lab, the results are quite different.

Before going to the individual laboratories, Figure 4 (not available) is an example of a bias plot illustrating all of the results from all of the laboratories for all of the shipped samples. The CD4+ count obtained by the predicate method is plotted on the x-axis, and the difference between the 2 methods (CD4 by BD-TC minus CD4 by PM) for each individual CD4+ result is plotted on the y-axis. At a glance it appears that most of

the results are evenly spread out around the zero difference line. This is what you want you want to see when you are going to implement a new technology.

When you look at the individual laboratories you get a much different picture. In subtracting the CD4 by the PM from the CD4 by the True-Count® method, the medians of the differences were greater than zero in three of five laboratories indicating that the TrueCount® CD4+ values were higher than the PM CD4+ counts in 3 of the 5 laboratories. The differences between the laboratories was in large part determined by the type of hematology instrumentation being used to generate the absolute lymphocyte count. Figure 5 is an example of a bias plot where the CD4+ counts generated by the BD-TC were greater than the CD4+ counts generated by the PM. When one subtracts the CD4+ by PM from the CD4+ by BD-TC, most of the values are positive, and again, there is a systematic bias, i.e., as the CD4+ count increases, the difference between the results increases. Figure 6 is an example of just the opposite result in a laboratory where the CD4+ counts generated by the PM were higher than the CD4+ counts generated by BD-TC. Therefore when subtracting the PM from the BD-TC, a negative result is obtained and is graphically represented with most of the results below the zero difference line. Again, there is a systematic bias.

Table 5 is a summary of the accuracy assessment in the Beckman/Coulter study, presenting the medians of the differences between the PM and the TetraOne® system (CD4 TetraOne minus CD4 PM). This study is not complete and the table represents only a preliminary assessment. In four of the five labs, the predicate method generated higher CD4+ counts than the Tetra-One® method. Again, there seems to be different biases in different laboratories, but overall, the TetraOne® system generated CD4+ values which were lower than the CD4+ values generated by the predicate method.

For the evaluation of intralaboratory variability, each laboratory performed CD4+ counts on replicate samples by both the predicate method and the new technology. Every laboratory obtained the statistically appropriate pre-determined number of local donors and split each donor sample into eight replicates, each treated as individual specimens. We then calculated the mean and standard deviation of the CD4+ replicates for each patient by each method. From these results we were able to calculate a percent CV by both methods on each patient in each laboratory. Every single sample has a percent CV for each of the methods.

For the inter-laboratory variability assessment, each of the laboratories in each study received an aliquot of the same samples which were shipped from a central location (FAST Systems Inc), i.e., five laboratories analyzed the same shipped specimen by the new single platform technology and by the predicate method. For each specimen, the mean CD4+ count, the standard deviation, and the percent CV for the predicate method and for the new technology were calculated. In order to assess the variability of the new technologies relative to the PM, subtract one from the other. If the variability (%CV) is the same for each of the methods then the difference would be zero and that is our null hypothesis for the precision analyses both within and between laboratories.

Table 6 summarizes the inter- or between laboratory variability results obtained with both the predicate method and the single platform technologies for each of the studies. The smaller font size represents the median interlaboratory %CV for the PM CD4+ counts in each of the three studies. The coefficient of variation for the PM ranges from 15-20% overall (all CD4+ categories), and from 19-30%, 13-18% and 11-22% in the < 200, 200-500 and >500 CD4 categories, respectively. These are relatively high interlaboratory variability values (%CVs). When performed by flow cytometry alone the CVs are around 5%. In all likelihood, it is the hematol-

ogy instrumentation which increases the variability in CD4+ counts between laboratories. The larger font size represents the interlaboratory %CV obtained using the single platform technologies. Overall in the Biometric® study we went from a CV of 15% with the predicate method down to 9%, in the BD study, we went from 16% down to 9%; we went from 19% down to 9.5% in the Beckman/Coulter study. The single platform systems have reduced the between laboratory variability significantly. Of note is the variability observed with the Imagn® in the group of <200 CD4+ T cells. This device actually counts the number of cells in approximately one &L of blood. If the patient has a CD4+ count <50, the error contributed simply by the counting error is significant. This is evident in the group of < 200 CD4+ counts where there was no improvement in the between laboratory precision in the Biometric® study. This was an issue three years ago when an increasing number of patients were being seen in the clinics with CD4+ counts below 100, and we were assuming the counts would keep decreasing in the population as a whole necessitating improved precision for low CD4+ counts. However the CD4+ counts are not going down, they are coming back up due to the remarkable improvements in antiretroviral therapy. In summary we observed a significant improvement in the between laboratory variability with the single platform technologies.

What do you buy by bringing these methods into your laboratory in terms of within laboratory variability? The answer is not very much. The within laboratory variability of absolute counting today is very good. Table 7 is a summary of the within laboratory variability of the three studies overall and in each of the three CD4+ groups. The smaller fonts represent the results obtained by the predicate method in each of the studies. The within laboratory variability of the predicate method overall is in most cases <10%. We saw a small but statistically significant improvement in some of the cases but, overall,

within laboratory variability of the predicate method is only slightly improved as compared with the new technologies.

The last aspect of the evaluation was "temporal fortitude", or how precise and accurate the new technology is on aged samples. We measured the CD4+ count on the sample when it was fresh (within six hours) and again at 24 hours using the same methodologies. We subtracted the CD4+ count obtained on the fresh specimen from the CD4 count obtained on the 24 hour old specimen. No change in the counts would yield a difference of zero. Table 8 is a summary of the median differences (CD4+ @ 24 hours minus CD4+ fresh). When the samples were analyzed by the predicate method, there were more CD4+ cells at 24 hours than there were at 6 hours. Unless the lymphocytes are dividing in the tube, this is probably not correct and is most likely due to the inclusion of non-lymphocytes in the lymphocyte gate during the automated differential on aged specimens. When measuring CD4+ counts in the aged samples by the single platform technologies, we observed the counts were slightly but significantly lower than the counts on the specimen when they were fresh. In summary, the single platform technologies showed a decrease in CD4+ counts at 24 hours compared to 6 hours, whereas the predicate method in two different studies showed an increase in CD4+ counts at 24 hours compared to CD4+ counts on fresh samples. Erroneous results on aged samples is a recognized phenomenon in the hematology laboratory.

Although the study was not designed to investigate the differences in precision due to differences in lymphocyte gating techniques, we observed improved precision in CD4+ counts by gating on lymphocytes by CD45 expression and light scatter compared to gating on lymphocytes by light scatter alone. Figure 7 illustrates the gating method used in the Coulter FlowCount® versus Coulter TetraOne® proto-

cols, and summarizes the intralaboratory variability. The only difference in these two methods is that with the FlowCount® protocol lymphocytes are gated on light scatter, and in the TetraOne® protocol lymphocytes are gated via fluorescence intensity of CD45. The intralaboratory %CVs for light scatter vs. CD45 gating is 12% versus 9%, 11% versus 9%, and 15% versus 10% for the <200, 200-500, and >500 CD4+ categories, respectively. There was a statistically significant improvement in precision by using CD45 gating as opposed to light scatter gating.

SUMMARY

In summary, evaluation of the accuracy of the new technologies for all of the laboratories collectively appears to be excellent, however this is not the way to evaluate these technologies. Examining individual laboratories is required and in doing so we saw bias. Each laboratory has a bias which is due predominantly to the type of hematology instrumentation used. Ten of the 15 laboratories generated higher CD4+ values by the predicate method than the values obtained by the single platform technology.

The between laboratory precision was significantly improved with all of the single platform technologies with the exception of the Biometric Imagn® when analyzing samples with CD4+ counts <200 cells/&L. This result was expected due to the counting error when counting low frequency events. The within laboratory variability was only slightly improved, and lymphocyte gating based on CD45 fluorescence significantly improves the precision of CD4+ counts as compared to light scatter gating.

With respect to "temporal fortitude", we observed some interesting things, and can infer some others. We saw small but significant differences on aged samples as compared to fresh samples with the single platform technologies. It is interesting that the CD4+ values obtained by

the predicate method increase after 24 hours. This is a clue that absolute lymphocyte counts generated on aged specimens are probably not accurate with most hematology instruments. The CD4+ values obtained by the new single platform technologies decreased after 24 hours, and this was somewhat expected. These were small but significant decreases. There is a potential problem with this analysis because in our studies, any sample that didn't generate a value at 24 hours was excluded. This happened most often with the predicate method because the sample failed our quality control criteria. We are therefore underestimating how many problems there were on aged samples when analyzed by the predicate method.

In summary, the single platform technologies generate reproducible CD4+ results both within laboratories, but more impressively represent a significant improvement in between laboratory variability as compared to the current predicate methods. With respect to accuracy, the new single platform technologies generate biased results which appear to depend on the type of hematology analyzer utilized. With respect to "temporal fortitude", the single platform technologies appear to generate slightly but significantly

lower CD4+ values at 24 hours compared to the values generated on fresh samples. This is expected. The predicate methods, on the other hand, generated higher CD4+ counts at 24 hours compared to fresh counts. This is erroneous and probably due to the inclusion of non-lymphocytes in the absolute lymphocyte count generated by the hematology analyzers on aged samples.

Acknowledgments

*The members of the NTEG include Rebecca Gelman, Cindy Wilkening and John Spritzler, statisticians from the Harvard Medical School, Dr. Nicholson from the CDC, Frank Mandy from Health Canada, Stephen Douglas from CHOPs in Philadelphia, John Schmitz from University of North Carolina, Eileen Bessent from UCSD, Albert Donnenberg from the University of Pittsburgh, Carol Bick from Indiana University, Keith Reiman from the Harvard Medical School and Daniella Livnat and Jonathan Kagan the NIAID representatives. I want to acknowledge these individuals because they have been involved since the beginning and have been instrumental in the planning, and execution of these evaluations.

Figure 1a

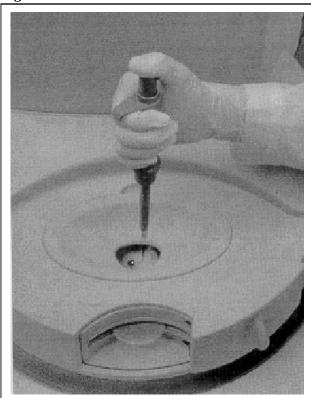


Figure 1b

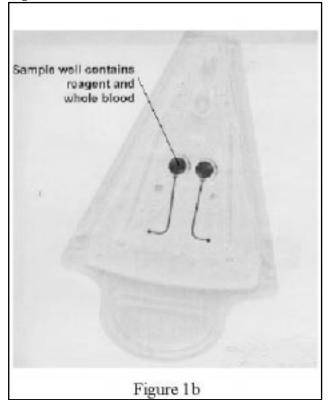


Figure 2

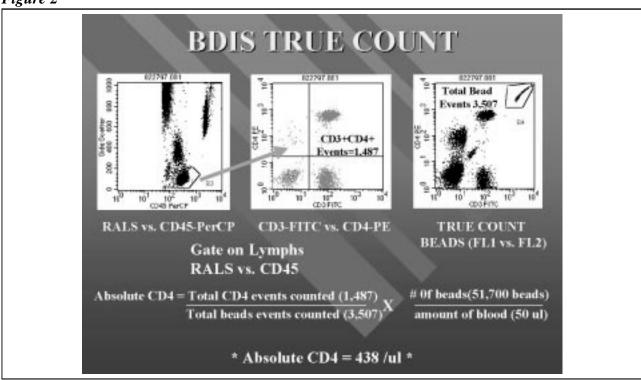


Figure 3

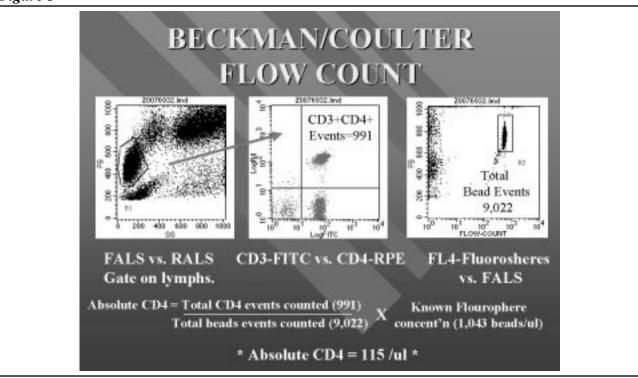


Table 1

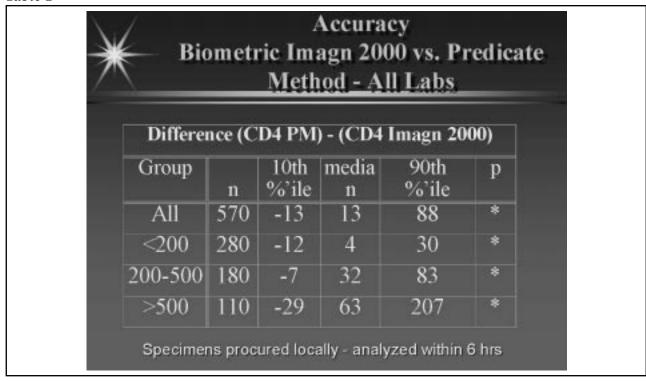


Table 2

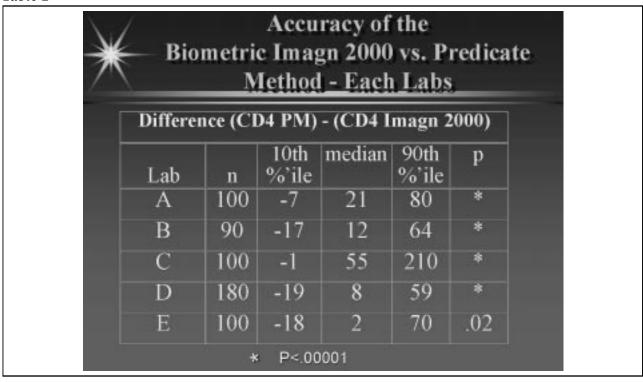
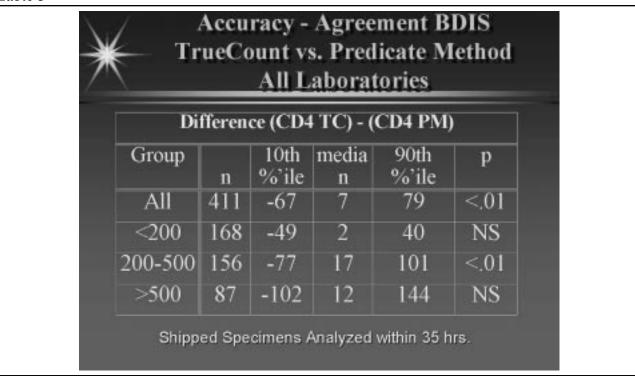
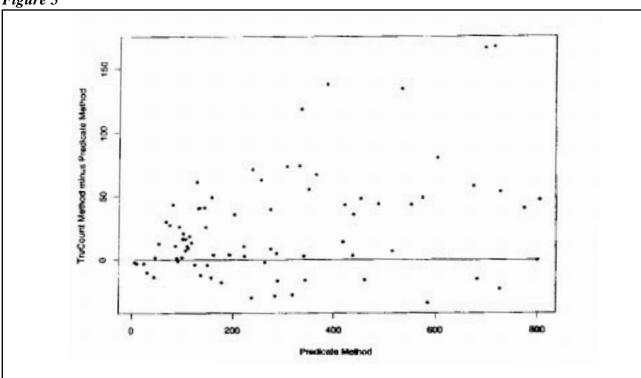


Table 3









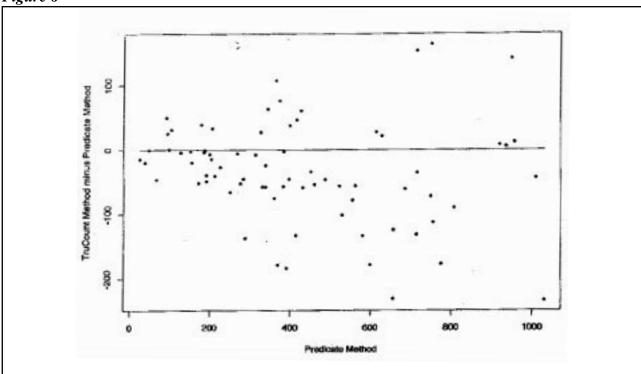


Table 4

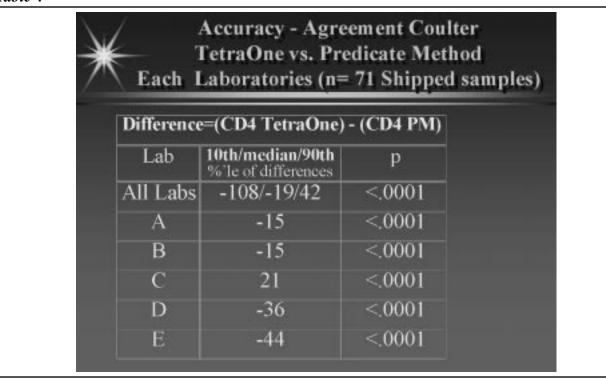


Table 5

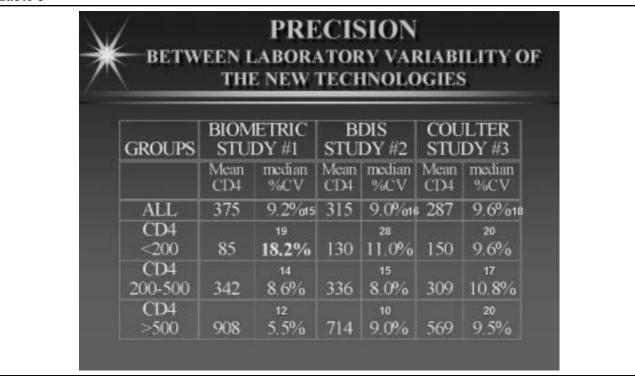


Table 6

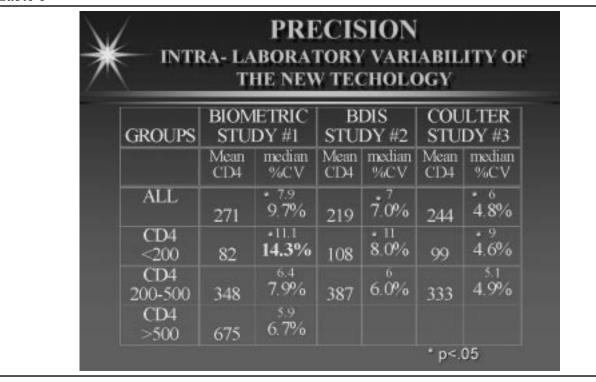


Table 7

